# Convergence rates of the continuous regularized Gauss-Newton method *

Barbara Kaltenbacher
SFB013 “Numerical and Symbolic Scientific Computing”
University of Linz
4040 Linz, Austria
barbara@sfb013.uni-linz.ac.at

Andreas Neubauer
Industial Mathematics Institute
University of Linz
4040 Linz, Austria
neubauer@indmath.uni-linz.ac.at

Alexander G. Ramm†
Department of Mathematics
Kansas State University
Manhattan, Kansas 66506-2602, USA
ramm@math.ksu.edu

**Abstract.** In this paper a convergence proof is given for the continuous analog of the Gauss-Newton method for nonlinear ill-posed operator equations and convergence rates are obtained. Convergence for exact data is proved for nonmonotone operators under weaker source conditions than before. Moreover, nonlinear ill-posed problems with noisy data are considered and a priori and a posteriori stopping rules are proposed. These rules yield convergence of the regularized approximations to the exact solution as the noise level tends to zero. The convergence rates are optimal rates under the source conditions considered.

**Key words.** nonlinear ill-posed problems, operator equations, dynamical systems, differential inequalities, regularization, continuous Gauss-Newton method, convergence rates.

**AMS(MOS) 2000 Subject Classifications:** 65J15, 65J20, 58C15, 47H17

---

*Supported by the Austrian Science Foundation Fonds in the Special Research Program SFB F013 and under grant T7-TEC

†visiting the University of Linz; support by SFB F013 is gratefully acknowledged

## 1. INTRODUCTION AND MAIN RESULTS

Consider the nonlinear operator equation

$$F(u) = f \tag{1.1}$$

where $F : H_1 \to H_2$ is a nonlinear operator between real Hilbert spaces $H_1$ and $H_2$. Assume that (1.1) is (not necessarily uniquely) solvable, i.e., there exists a $y \in H_1$ such that

$$F(y) = f\,. \tag{1.2}$$

We are interested in ill-posed probems (1.1) when $u$ does not depend in a stable way on the data $f$ and the given data $f_\delta$ are the exact data $f$ contaminated by noise, so that $f_\delta$ is given such that

$$\|f - f_\delta\| \le \delta\,. \tag{1.3}$$

Our analysis is local: an initial guess $u_0$ is assumed to be sufficiently close to $y$, i.e., $u_0 \in B(y, \rho)$ for a suitable $\rho > 0$.

For the stable solution of (1.1), we consider the continuous analog of the Gauss-Newton method, namely

$$\begin{aligned}\dot{u}(t) = -[F'(u(t))^* F'(u(t)) + \varepsilon(t) I]^{-1}\, &[F'(u(t))^*(F(u(t)) - f)\\ &+ \varepsilon(t)(u(t) - u_0)], \quad t > 0, \\ u(0) = u_0\,. \end{aligned} \tag{1.4}$$

Here, $\varepsilon$ is a continuously differentiable strictly positive function with strictly positive values, decreasing strictly monotonically to zero as $t \to \infty$,

$$\varepsilon : [0, \infty) \to (0, \infty) \in C^1([0, \infty)),\ \varepsilon \searrow 0 \text{ as } t \to \infty,\ \dot{\varepsilon} := \frac{d\varepsilon}{dt} < 0, \tag{1.5}$$

and

$$\frac{|\dot{\varepsilon}(t)|}{\varepsilon(t)} \le c_\varepsilon < 1 \tag{1.6}$$

with some sufficiently small constant $c_\varepsilon > 0$ (see assumptions (1.14) and (1.16) in the convergence theorems below). Note that these conditions allow for not only polynomially but even exponentially decaying $\varepsilon$, which enables a fast approximation of the $F'(u(t))^{-1}$ by the operator $[F'(u(t))^* F'(u(t)) + \varepsilon(t) I]^{-1} F'(u(t))^*$ and is therefore of importance for the speed of convergence of the method (see the error estimate (1.17) below).

In [1]–[3] and [13]–[17] a general approach to solving linear and nonlinear ill-posed problems is developed. This approach consists of finding a Cauchy problem (a dynamical system)

$$\dot{u}(t) = \Phi(t, u(t)), \quad t > 0, \quad u(0) = u_0,$$

such that the following three conditions hold:

(i) this problem has a unique global solution $u(t)$ for any $u_0 \in B(y, \rho)$,

(ii) this solution has a limit as $t \to \infty$, $\lim_{t\to\infty} u(t) = u(\infty)$

and

(iii) this limit solves equation (1.1): $F(u(\infty)) = f$.

Examples of $\Phi$ for which conditions (i) – (iii) hold are given in [13] for a wide class of linear ill-posed problems, in [1]–[3], [14], [16], [17] for a wide class of nonlinear ill-posed problems with monotone $F$, and for nonmonotone $F$, under the assumption that $F$ is locally smooth (twice Fréchet differentiable) and satisfies a source type assumption:

$$u_0 - y = (F'(y)^* F'(y))^\nu v, \tag{1.7}$$

with some $v \in H_1$ and with $\nu \geq \frac{1}{2}$ (see, e.g., Lemma 4.6 in [3] for the case $\frac{1}{2} \leq \nu \leq 1$).

The aim of this paper is to prove convergence properties (i), (ii) and (iii) for $\Phi(t, u)$ chosen in (1.4), that is, for the continuous analog of the Gauss-Newton method, without monotonicity assumption on $F$ and with weaker source conditions (1.7) for some $v \in H_1$ and some $0 < \nu < \frac{1}{2}$ or

$$u_0 - y = (-\ln(\omega F'(y)^* F'(y)))^{-p} v \tag{1.8}$$

for some $v \in H_1$ and some $p > 0$, (where $0 < \omega \leq \frac{1}{e\|F'(y)\|^2}$ is a constant scaling factor chosen so that the argument of the function $(-\ln(\cdot))^{-p}$ is smaller than $\frac{1}{e}$ and hence remains bounded away from the singularity at 0, which is used in the spectral representation of the operator in (1.8)). We also consider the case when no regularity conditions on $u_0 - y$ are given, assuming only that

$$u_0 - y \in N(F'(y))^\perp, \tag{1.9}$$

which is the minimal assumption to get convergence only, without a rate. Moreover, the case of noisy data and corresponding stopping time rules are considered. They were not discussed in [3].

Let us explain, why this question is important: Typically in inverse problems the operators $F'(u)$ and $F'(u)^*$ are smoothing. The source conditions (1.7) and (1.8) are therefore smoothness assumptions on the initial error $u_0 - y$. Since the function $\lambda \mapsto \lambda^\nu$ decays faster than $\lambda \mapsto (-\ln(\lambda))^{-p}$ as $\lambda \to 0$, condition (1.8) for some $p > 0$ is weaker than (1.7) for $\nu > 0$, and (1.7) is the weaker, the smaller $\nu$ is; $\nu = 0$ means that no smoothness assumption is made. In exponentially ill-posed problems (e.g., inverse scattering if the potential is compactly supported, and other inverse problems), where the operator $F'(u)$ is infinitely smoothing, (1.7) with $\nu > 0$ would mean that the nonsmooth part of the solution has to be known exactly. Such an assumption is too strong, while condition (1.8) requires $u_0 - y$ only to be in some Sobolev space of finite order and therefore is more realistic (see [9, 5, 10]). For moderately ill-posed problems such as some parameter identification problems, where the linearized forward operator $F'(u)$ is smoothing of finite order, condition (1.7) with small $\nu$ is more likely to hold than one with $\nu \geq \frac{1}{2}$, e.g., if solutions have jumps whose precise locations are not known.

Our result consists of two parts. The first part contains a convergence proof for the problem with exact data. The second part treats the case of noisy data. For noisy data with a fixed noise level due to the ill-posedness it is impossible, in general, to prove convergence to the solution, but if the noise level tends to zero one can prove existence of a moment of time $t_\delta$ such that $||u(t_\delta) - y|| \to 0$ as $\delta \to 0$. The rule for choosing $t_\delta$ we call the stopping rule.

The first part of our results is the basic one because there is a general principle (see [12]) which says, roughly speaking, that if one can construct a method for solving an ill-posed problem with exact data, then one can modify this method to get a stable approximation to the solution of the ill-posed problem when the data are noisy. More precisely, suppose $F(y) = f$, one knows a family of operators $R_n$, such that $||R_n(f) - y|| \to 0$ as $n \to \infty$ and for any fixed $n$ the operator $R_n$ is continuous, and let $f_\delta$, the noisy data, be given, such that $||f - f_\delta|| \leq \delta$. Denote $u_\delta := R_n(f_\delta)$. Then

$$||y - u_\delta|| \leq ||y - R_n(f)|| + ||R_n(f) - R_n(f_\delta)|| \leq a(n, f) + b(n, \delta)\,,$$

where $a(n, f) \to 0$ as $n \to \infty$ by the definition of $R_n$ and $b(n, \delta) \to 0$ as $\delta \to 0$, $n$ being fixed, by the continuity of $R_n$ with a fixed $n$. If the problem is ill-posed and $f_\delta$ is not in the range of $F$, then $b(n,\delta) \to \infty$ as $n \to \infty$ for a fixed $\delta$. Therefore, the problem

$$a(n,f) + b(n,\delta) = \min$$

has a solution $n(\delta)$, $n(\delta) \to \infty$ as $\delta \to 0$, and

$$E(\delta) := a(n(\delta),f) + b(n(\delta),\delta) \to 0 \text{ as } \delta \to 0\,.$$

Thus, a stable approximation to the solution $y$ is given by the formula $u_\delta = R_{n(\delta)} f_\delta$ and the error of this approximation is: $||y - u_\delta|| \leq E(\delta)$. This is an example of the usage of the notion of regularizing algorithm.

We assume that $F$ is Fréchet differentiable with uniformly bounded derivative

$$\|F'(u)\| \leq C_F \quad \text{for all } u, \bar{u} \in B(y,\rho) \tag{1.10}$$

and either

$$\|F'(\bar{u}) - F'(u)\| \leq L\|\bar{u} - u\| \quad \text{for all } u, \bar{u} \in B(y,\rho) \tag{1.11}$$

or

$$\begin{matrix} F'(\bar{u}) = F'(u)R(\bar{u},u) \quad \text{and} \\ \|R(\bar{u},u) - I\| \leq C_R\|\bar{u} - u\| \end{matrix} \quad \text{for all } u, \bar{u} \in B(y,\rho) \tag{1.12}$$

with some linear operators $R(\bar{u},u) : H_1 \to H_1$, or

$$\begin{matrix} F'(\bar{u}) = R(\bar{u},u)F'(u) \quad \text{and} \\ \|R(\bar{u},u) - I\| \leq c_R < 1 \end{matrix} \quad \text{for all } u, \bar{u} \in B(y,\rho) \tag{1.13}$$

with some linear operators $R(\bar{u},u) : H_2 \to H_2$. Note that conditions (1.12) or (1.13) are more specific and often harder to verify for concrete applications than (1.11). Assumptions of this type allow one to prove convergence of regularization methods for nonlinear problems when no monotonicity can be used and only (1.7) with $\nu < \frac{1}{2}$, or (1.8), or just (1.9) holds (see [4]–[11]). For some examples of nonlinear inverse ill-posed problems satisfying (1.12), or (1.13), see [11] and [7], respectively.

Our main result for exact data is the following

**Theorem 1.1.** *Let the data be exactly given ($\delta = 0$ in (1.3)) and let $u_0 - y \in N(F'(y))^\perp \cap B(y,\rho)$. Assume that conditions (1.5) and (1.6) on $\varepsilon$ and (1.10) on $F$ are satisfied and that one of the following cases occurs:*

(i) *$F$ satisfies (1.11), (1.7) holds with $\nu \geq \frac{1}{2}$ and assume*

$$C_2 > 0\,, \quad 4C_1C_3 < C_2^2\,,$$

$$\frac{\|u_0 - y\|}{\|\varepsilon(0)[F'(y)^*F'(y) + \varepsilon(0)I]^{-1}(u_0 - y)\|} < \frac{C_2 + \sqrt{C_2^2 - 4C_1C_3}}{2C_3}, \tag{1.14}$$

$$\text{and} \quad \tfrac{C_1}{C_2}\|\varepsilon(0)[F'(y)^*F'(y) + \varepsilon(0)I]^{-1}(u_0 - y)\| < \rho\,,$$

*where*

$$C_1 := 1\,, \quad C_2 := 1 - \tfrac{5L}{4}\|F'(y)\|^{2\nu-1}\|v\| - c_\varepsilon\,,$$
$$C_3 := \tfrac{L}{4}\nu^\nu(1-\nu)^{1-\nu}\varepsilon(0)^{\nu-\frac{1}{2}}\|v\|\,.$$

(ii) *$F$ satisfies (1.12), (1.7) with $0 \leq \nu \leq \frac{1}{2}$ or (1.8) holds, and (1.14) with*

$$C_1 := 1\,, \quad C_2 := 1 - 2C_R\|u_0 - y\| - c_\varepsilon\,, \quad C_3 := \tfrac{1}{2}C_RC_0\|v\|\,,$$

$$C_0 = \begin{cases} \nu^\nu(1-\nu)^{1-\nu}\varepsilon(0)^\nu & \text{if } (1.7) \text{ holds} \\ \gamma_1(p) & \text{if } (1.8) \text{ holds} \end{cases} \tag{1.15}$$

*with the constant $\gamma_1(p)$ defined in Lemma 2.1 below.*

(iii) *$F$ satisfies (1.13), (1.7) with $0 \leq \nu \leq \frac{1}{2}$ or (1.8) holds, and*

$$C_2 > 0\,, \qquad \tilde{C}_2 > 0\,, \quad \text{and} \quad \frac{C_1}{C_2}r(0) < \rho \tag{1.16}$$

*with*

$$C_1 := c_R(1 + c_R)\max\{\tfrac{\tilde{C}_1}{\tilde{C}_2}, \tfrac{\|F'(y)(u_0-y)\|}{\|\varepsilon(0)F'(y)[F'(y)^*F'(y)+\varepsilon(0)I]^{-1}(u_0-y)\|}\}$$
$$+ \max\{1, c_R(1 + c_R)\}$$
$$C_2 := 1 - 2c_\varepsilon\,, \quad \tilde{C}_1 := 1 + 2c_R(1 + c_R)^2\,,$$
$$\tilde{C}_2 := 1 - c_R(1 + c_R)^2 - c_\varepsilon\,.$$

*Then for all $t \geq 0$, the solution $u(t)$ to (1.4) exists, is unique, and lies in $B(y, \rho)$, and*

$$u(t) \to y \text{ as } t \to \infty\,.$$

*Moreover,*

$$\|u(t)-y\| \le C \begin{cases} \varepsilon(t) & \text{if (1.7) holds with } \nu \ge 1 \\ \varepsilon(t)^\nu & \text{if (1.7) holds with } 0<\nu<1 \\ \left(-\ln\left(\frac{1}{e\varepsilon(0)}\varepsilon(t)\right)\right)^{-p} & \text{if (1.8) holds} \end{cases} \tag{1.17}$$

*for some constant* $C > 0$*.*

**Remark 1.2.** Conditions (1.14) or (1.16), respectively, will be called closeness conditions on the initial approximation and the data because they can always be satisfied if $\|u_0 - y\|$, $c_R$, and $\|v\|$ are sufficiently small.

The second part of our results consists of considering problem (1.4) with noisy data:

$$\begin{aligned} \dot{u}_\delta(t) &= -(F'(u_\delta(t))^*F'(u_\delta(t)) + \varepsilon(t)I)^{-1} \\ &\qquad [F'(u_\delta(t))^*(F(u_\delta(t)) - f_\delta) + \varepsilon(t)(u_\delta(t) - u_0)], \quad t > 0, \\ u_\delta(0) &= u_0\,. \end{aligned} \tag{1.18}$$

In this case one finds a stopping time $t_\delta$, that is, a stopping rule, such that:

$$u_\delta(t_\delta) \to y \quad \text{as} \quad \delta \to 0\,,$$

with optimal rates under source type assumptions. As mentioned above, Theorem 1.1 guarantees existence of such a stopping rule. To give a flavor of a possible concrete order optimal choice of $t_\delta$, we do an analysis of the propagation of the data noise (see Lemmas 2.2 – 2.4 below), and derive an a priori stopping rule:

In order to get convergence when no source type assumptions hold it suffices that $t_\delta$ satisfies

$$t_\delta \to \infty \quad \text{and} \frac{\delta}{\sqrt{\varepsilon(t_\delta)}} \to 0 \quad \text{as} \quad \delta \to 0. \tag{1.19}$$

To obtain optimal rates in case of source type assumptions $t_\delta$ has to be chosen as solution of the equation

$$\varepsilon(t_\delta)^{\nu+\frac{1}{2}} = \tau\delta \tag{1.20}$$

if (1.7) holds or

$$\varepsilon(t_\delta) = \tau\delta \tag{1.21}$$

if (1.8) holds, with a sufficiently large constant $\tau > 0$. Note that this corresponds to the order optimal regularization parameter choice in Philips-Tikhonov regularization (see, e.g., [6], [10]).

**Corollary 1.3.** *Let the assumptions of Theorem 1.1 be satisfied, with the exception that the data are contaminated with noise, with a noise level $\delta$, according to (1.3) and the process is stopped according to the stopping rule (1.19)-(1.21) with some sufficiently large constant $\tau$.*

*Then for all $0 \le t \le t_\delta$, $u_\delta(t)$, the solution to (1.18) lies in $B(y, \rho)$,*

$$u_\delta(t_\delta) \to y \quad \text{as} \quad \delta \to 0, \tag{1.22}$$

*and*

$$\|u_\delta(t_\delta) - y\| = \begin{cases} O(\delta^{\frac{2}{3}}) & \text{if (1.7) holds with } \nu \ge 1 \\ O(\delta^{\frac{2\nu}{\nu+1}}) & \text{if (1.7) holds with } 0 < \nu < 1 \\ O((-\ln(\delta))^{-p}) & \text{if (1.8) holds.} \end{cases} \tag{1.23}$$

For obtaining optimal convergence rates (1.23) the stopping rules (1.20) and (1.21) obviously need some a priori information on the type of source condition and, in case of (1.7), on the exponent $\nu$. While it is usually known whether logarithmic or Hölder type conditions, respectively are likely to hold, (namely for exponentially or moderately ill-posed problems, respectively,) explicit knowledge of $\nu$ in (1.7) is in practice often not available. To have a practically applicable optimal stopping rule also for Hölder type source conditions (1.7), consider the following generalization of the discrepancy principle: with some sufficiently large constant $\tau > 1$, the stopping time is chosen by the formula:

$$\|F(u_\delta(t_\delta)) - f_\delta\| = \tau\delta\,, \tag{1.24}$$

and we assume that

$$\tau\delta < \|F(u_\delta(t)) - f_\delta\| \text{ for all times } t < t_\delta\,, \tag{1.25}$$

i.e., $t_\delta$ is the first moment $t$, at which the discrepancy is equal to $\tau\delta$. If

$$\|F(u_0) - f_\delta\| > \tau\delta, \tag{1.26}$$

then formulas (1.24) and (1.25) determine uniquely $t_\delta > 0$. Condition (1.26) means that we impose a lower bound on the signal-to noise ratio $\frac{\|F(u_0)-f_\delta\|}{\delta}$.

For a possibly large but fixed constant $\tau > 0$, this can be satisfied if the noise is sufficiently small compared to the initial residual. Note that if (1.26) is not satisfied, i.e., if $\|F(u_0) - f_\delta\|$ is not significantly larger than $\delta$, or in other words, the initial residual is already of the order of magnitude of the noise level, one cannot expect to get a significantly better approximation of the solution from the given data than the initial guess $u_0$ itself.

**Theorem 1.4.** *Let the data $f_\delta$ satisfy (1.3) and (1.26) and let $u_0 - y \in N(F'(y))^\perp \cap B(y,\rho)$. Assume that conditions (1.5) and (1.6) on $\varepsilon$, (1.10) and (1.13) on $F$ are satisfied, and (1.7) with*

$$0 \le \nu \le \frac{1}{2},$$

$$C_2 > 0\,, \qquad \overline{\tilde{C}}_2 > 0\,, \quad \textit{and} \quad \frac{\overline{C}_1}{C_2} r(0) < \rho \tag{1.27}$$

*where*

$$\begin{aligned}
&\overline{C}_1 := (c_R + \tfrac{1}{2(\tau-1)})(1+c_R)\max\{\tfrac{\tilde{C}_1}{\overline{\tilde{C}}_2}, \tfrac{\|\mathcal{A}(u_0-y)\|}{\tilde{r}(0)}\} + \max\{1, c_R(1+c_R)\},\\
&C_2 := 1 - 2c_\varepsilon,\\
&\tilde{C}_1 := 1 + 2c_R(1+c_R)^2,\\
&\overline{\tilde{C}}_2 := 1 - c_R(1+c_R)^2 - c_\varepsilon - (1+c_R)^2\tfrac{1}{\tau-1}.
\end{aligned}$$

*Then for all $0 \le t \le t_\delta$, $u_\delta(t)$ the solution to (1.18) lies in $B(y,\rho)$,*

$$u_\delta(t_\delta) \to y \quad \textit{as} \quad \delta \to 0\,, \tag{1.28}$$

*and*

$$\|u_\delta(t_\delta) - y\| \le C\delta^{\frac{2\nu}{2\nu+1}} \tag{1.29}$$

*for some constant $C > 0$ independent of $\delta$.*

Convergence (1.28) follows from (1.29) if $\nu > 0$ but also holds for $\nu = 0$, in which case no rate of convergence can be obtained.

Note that for the case $\nu \ge \frac{1}{2}$ and exact data we basically repeat the arguments used in [3]. This is done for completeness of the presentation and because we use these arguments in Corollary 1.3 in the case of noisy data, not considered in [3].

The proofs are based on a combination of methods from [1]–[3] and [13]–[17] with ideas similar to those used in [4] for the convergence analysis of the iteratively regularized Gauss-Newton method. More precisely, we derive differential inequalities for the function

$$\psi_\delta(t) = \frac{\|u_\delta(t) - y\|}{r(t)}$$

where

$$r(t) = \|\varepsilon(t)(F'(y)^*F'(y) + \varepsilon(t)I)^{-1}(u_0 - y)\| \,,$$

or for the function

$$\tilde{\psi}_\delta(t) = \frac{\|F'(y)(u_\delta(t) - y)\|}{\tilde{r}(t)} \,,$$

where

$$\tilde{r}(t) = \|\varepsilon(t)F'(y)(F'(y)^*F'(y) + \varepsilon(t)I)^{-1}(u_0 - y)\| \,.$$

The terms $r(t)$, $\tilde{r}(t)$ go to zero as $t \to \infty$ at a rate determined by the source condition (1.7) or (1.8), or in general go to zero, as $t \to \infty$, arbitrarily slowly if only (1.9) is assumed to hold. From this we conclude, under some closeness conditions (see the assumptions (1.14) or (1.16)), uniform boundedness of $\psi$ for all times in the case of exact data (i.e., in the case $\delta = 0$, $f_\delta = f$, $u_\delta = u$, $\psi_\delta = \psi$) and up to the stopping time $t_\delta$ in the situation of noisy data, respectively, which gives us the stated convergence results.

## 2. AUXILIARY RESULTS

**Lemma 2.1.** *For any bounded linear operator $A : H_1 \to H_2$, $0 < \varepsilon_0$, $0 \le \mu \le 1$ (setting $0^0 := 1$) one has for all $\varepsilon \le \varepsilon_0$:*

$$\|\varepsilon(A^*A + \varepsilon I)^{-1}(A^*A)^\mu\| \le \mu^\mu(1-\mu)^{1-\mu}\,\varepsilon^\mu \,, \tag{2.1}$$

$$\|\varepsilon(A^*A + \varepsilon I)^{-1}(-\ln(\tfrac{1}{e\|A\|^2}A^*A))^{-p}\| \le \gamma_1(p)(-\ln(\tfrac{1}{e\varepsilon_0}\varepsilon(t))^{-p} \,, \tag{2.2}$$

$$\|\varepsilon A(A^*A + \varepsilon I)^{-1}(-\ln(\tfrac{1}{e\|A\|^2}A^*A))^{-p}\| \le \gamma_2(p)\sqrt{\varepsilon}(-\ln(\tfrac{1}{e\varepsilon_0}\varepsilon(t))^{-p} \,, \tag{2.3}$$

*with some constants $\gamma_1(p), \gamma_2(p)$. Moreover, for all $w \in H_1$ and all $v \in N(A)^\perp$*

$$\|\varepsilon(A^*A + \varepsilon I)^{-1}v\| \to 0 \,, \quad \|\sqrt{\varepsilon}A^*(A^*A + \varepsilon I)^{-1}w\| \to 0 \text{ as } \varepsilon \to 0 \tag{2.4}$$

*and conversely*

$$\|\varepsilon(A^*A + \varepsilon I)^{-1}w\| \to 0 \Rightarrow (\varepsilon \to 0 \text{ or } w \in N(A)) \tag{2.5}$$

**Proof of Lemma 2.1.** Let $E_\lambda$ be the resolution of the identity corresponding to the selfadjoint operator $A^*A$ and denote $m := \|A^*A\|$. Then, for any $w \in H_1$ one has

$$\|\varepsilon(A^*A+\varepsilon I)^{-1}(A^*A)^\mu w\|^2 = \int_0^{m+0} \left(\frac{\varepsilon}{\varepsilon+\lambda}\right)^2 \lambda^{2\mu}\, d(E_\lambda w, w)$$

where $(\ ,\ )$ is the inner product in $H_1$, and

$$\sup_{\lambda\in[0,m]} \frac{\varepsilon}{\varepsilon+\lambda}\lambda^\mu \le \mu^\mu(1-\mu)^{1-\mu}\varepsilon^\mu\,,$$

from which (2.1) follows. Similarly one proves (2.2) and (2.3) (cf. [10]). Note that

$$\frac{\varepsilon}{\varepsilon+\lambda}\begin{cases}\to 0 & \text{for } \lambda>0\\ =1 & \text{for } \lambda=0\end{cases}\,,\qquad \frac{\varepsilon\lambda}{(\varepsilon+\lambda)^2}\to 0 \quad \text{for all}\quad \lambda\ge 0 \quad \text{as}\quad \varepsilon\to 0\,.$$

Thus, using the formulas

$$A = U(A^*A)^{\frac{1}{2}}\,,$$

where $U : R(A^*A) \to R(A)$ is a partial isometry,

$$\|\varepsilon(A^*A+\varepsilon I)^{-1}v\|^2 = \int_0^{m+0}\left(\frac{\varepsilon}{\varepsilon+\lambda}\right)^2\, d(E_\lambda v, v)\,,$$

$$\|\sqrt{\varepsilon}A(A^*A+\varepsilon I)^{-1}w\|^2 = \int_0^{m+0}\frac{\varepsilon\lambda}{(\varepsilon+\lambda)^2}\, d(E_\lambda w, w)$$

and the assumption $v\in N(A)^\perp$, one gets (2.4). To show (2.5), observe that

$$\begin{aligned}&\|\varepsilon(A^*A+\varepsilon I)^{-1}w\|^2 = \int_0^{m+0}\left(\frac{\varepsilon}{\varepsilon+\lambda}\right)^2\, d(E_\lambda w, w)\\ &\ge \left(\frac{\varepsilon}{\varepsilon+m}\right)^2\int_0^{m+0} d\|E_\lambda w\|^2 = \left(\frac{\varepsilon}{\varepsilon+m}\right)^2 \|\operatorname{Proj}_{N(A)^\perp} w\|^2\end{aligned}$$

and the function $\varepsilon\mapsto\frac{\varepsilon}{\varepsilon+m}$ is strictly positive outside zero and strictly monotonically increasing.

Lemma 2.1 is proved. $\square$

The following Lemmas 2.2 – 2.4 also imply the respective differential inequalities in the case of exact data, i.e., with $\delta = 0$, for

$$\psi(t) := \frac{\|u(t) - y\|}{r(t)}\,, \qquad \tilde{\psi}(t) := \frac{\|F'(y)(u(t) - y)\|}{\tilde{r}(t)}\,,$$

and $u(t)$ the solution of (1.4)

Lemmas 2.2 – 2.4 contain as an assumption the statements about existence and uniqueness of $u(t)$ or $u_\delta(t)$ and the inclusion $u(t) \in B(y, \rho)$ or $u_\delta(t) \in B(y, \rho)$. In fact, local uniqueness and existence of $u$ and $u_\delta$ follow from the smoothness of the operator $\Phi$ on the right-hand side of (1.4) and (1.18), the inclusion $u(t) \in B(y, \rho)$, for all $t > 0$ is proved in Theorem 1.1, and the inclusion $u_\delta(t) \in B(y, \rho)$ for $t \le t_\delta$ is proved in Corollary 1.3 and Theorem 1.4.

**Lemma 2.2.** *Assume that conditions (1.5) and (1.6) on $\varepsilon$ and (1.10) and (1.11) on $F$ are satisfied, and that $u_\delta(t)$ the solution to (1.18) exists, is unique, and lies in $B(y, \rho)$. Moreover, let the source condition (1.7) with $\frac{1}{2} \le \nu \le 1$ hold. Then the following differential inequality holds*

$$\begin{aligned} \dot{\psi}_\delta(t) \le &-(1 - \frac{5L}{4}\|F'(y)\|^{2\nu-1}\|v\| - c_\varepsilon)\psi_\delta(t) \\ &+ \frac{L}{4}\nu^\nu(1-\nu)^{1-\nu}\varepsilon(0)^{\nu-\frac{1}{2}}\|v\|\,\psi_\delta(t)^2 \\ &+ 1 + \frac{\delta}{2\sqrt{\varepsilon(t)}r(t)} \end{aligned} \tag{2.6}$$

*with*

$$\psi_\delta(t) := \frac{\|u_\delta(t) - y\|}{r(t)}\,,$$

$$\begin{aligned} r(t) &:= \|\varepsilon(t)(F'(y)^*F'(y) + \varepsilon(t)I)^{-1}(u_0 - y)\| \\ &\le \overline{r}(t) := \nu^\nu(1-\nu)^{1-\nu}\varepsilon(t)^\nu\|v\|\,. \end{aligned} \tag{2.7}$$

**Proof of Lemma 2.2.** Denote

$$\begin{aligned} &A(t) := F'(u_\delta(t))\,, \quad \mathcal{A} := F'(y)\,, \quad e_\delta(t) := u_\delta(t) - y \\ &T_\varepsilon(u) = A(t)^*A(t) + \varepsilon(t)I\,, \quad \mathcal{T}_\varepsilon = \mathcal{A}^*\mathcal{A} + \varepsilon(t)I\,. \end{aligned}$$

It follows from (1.2) and (1.18) that

$$\begin{aligned}\dot{e_\delta}(t) &= -T_\varepsilon(u)^{-1}(A(t)^*(F(u_\delta(t)) - F(y) + f - f_\delta) + \varepsilon(t)e_\delta(t)\\ &\qquad + \varepsilon(t)(y-u_0))\\ &= -e_\delta(t) + T_\varepsilon(u)^{-1}A(t)^*(A(t)e_\delta(t) + F(y) - F(u_\delta(t)))\\ &\quad + T_\varepsilon(u)^{-1}A(t)^*(f_\delta - f) + \varepsilon(t)\mathcal{T}_\varepsilon^{-1}(u_0-y)\\ &\quad + \varepsilon(t)(T_\varepsilon(u)^{-1}([\mathcal{A}^* - A(t)^*]\mathcal{A}\\ &\qquad + A(t)^*[\mathcal{A} - A(t)]))\mathcal{T}_\varepsilon^{-1}(u_0-y)\,.\end{aligned} \tag{2.8}$$

Here the representation

$$T_\varepsilon(u)^{-1} - \mathcal{T}_\varepsilon^{-1} = T_\varepsilon(u)^{-1}([\mathcal{A}^* - A(t)^*]\mathcal{A} + A(t)^*[\mathcal{A} - A(t)])\mathcal{T}_\varepsilon^{-1}$$

was used.

Assumption (1.11) on $F$ implies

$$\|A(t)e_\delta(t) + F(y) - F(u_\delta(t))\| \le \frac{L}{2}\|e_\delta(t)\|^2\,, \tag{2.9}$$

$$\|\mathcal{A}^* - A(t)^*\| \le L\|e_\delta(t)\|\,, \quad \|\mathcal{A} - A(t)\| \le L\|e_\delta(t)\|\,. \tag{2.10}$$

From Lemma 2.1 one gets:

$$\|T_\varepsilon(u)^{-1}A(t)^*\| \le \frac{1}{2\sqrt{\varepsilon(t)}}\,, \quad \|T_\varepsilon(u)^{-1}\| \le \frac{1}{\varepsilon(t)}\,. \tag{2.11}$$

If $\nu \ge \frac{1}{2}$ in (1.7), then, using the fact that $R(\mathcal{A}^*) = R((\mathcal{A}^*\mathcal{A})^{\frac{1}{2}})$, one concludes that there exists a $\tilde{v} \in H_2$ such that

$$u_0 - y = \mathcal{A}^*\tilde{v}$$

and

$$\|\tilde{v}\| \le \|\mathcal{A}^*\mathcal{A}\|^{\nu-\frac{1}{2}}\|v\|\,.$$

Let us form the inner product $(e_\delta(t), \dot{e_\delta}(t))$. The relations (2.9), (2.10), (2.11), and Lemma 2.1 allow one to estimate each of the right hand side terms arising in (2.8):

$$|(T_\varepsilon(u)^{-1}A(t)^*(A(t)e_\delta(t) + F(y) - F(u_\delta(t))), e_\delta(t))| \le \frac{L}{4\sqrt{\varepsilon(t)}}\|e_\delta(t)\|^3\,,$$

$$|(T_\varepsilon(u)^{-1}A(t)^*(f_\delta - f), e_\delta(t))| \le \frac{\delta}{2\sqrt{\varepsilon(t)}}\|e_\delta(t)\|\,,$$

$$|(\varepsilon(t)\mathcal{T}_\varepsilon^{-1}(u_0 - y), e_\delta(t))| \le r(t)\,\|e_\delta(t)\|\,,$$

$$\begin{aligned}
&|(\varepsilon(t)T_\varepsilon(u)^{-1}[\mathcal{A}^* - A(t)^*]\mathcal{A}\mathcal{T}_\varepsilon^{-1}(u_0 - y), e_\delta(t))| \\
&\quad\le \|\varepsilon(t)T_\varepsilon(u)^{-1}\|\,\|\mathcal{A}^* - A(t)^*\|\,\|\mathcal{A}\mathcal{T}_\varepsilon^{-1}\mathcal{A}^*\tilde{v}\|\,\|e_\delta(t)\| \\
&\quad\le L\,\|\mathcal{A}^*\mathcal{A}\|^{\nu-\frac{1}{2}}\,\|v\|\,\|e_\delta(t)\|^2\,,
\end{aligned}$$

$$\begin{aligned}
&|(\varepsilon(t)T_\varepsilon(u)^{-1}A(t)^*[\mathcal{A} - A(t)]\mathcal{T}_\varepsilon^{-1}(u_0 - y), e_\delta(t))| \\
&\quad\le \|\varepsilon(t)T_\varepsilon(u)^{-1}A(t)^*\|\,\|\mathcal{A} - A(t)\|\,\|\mathcal{T}_\varepsilon^{-1}\mathcal{A}^*\tilde{v}\|\,\|e_\delta(t)\| \\
&\quad\le \tfrac{L}{4}\,\|\mathcal{A}^*\mathcal{A}\|^{\nu-\frac{1}{2}}\,\|v\|\,\|e_\delta(t)\|^2\,.
\end{aligned}$$

Since $H_1$ is a real Hilbert space, one gets:

$$\begin{aligned}
\tfrac{d}{dt}(\|e_\delta(t)\|^2) &= 2(\dot{e_\delta}(t), e_\delta(t)) \\
&\le 2\Big(-(1 - \tfrac{5L}{4}\,\|\mathcal{A}^*\mathcal{A}\|^{\nu-\frac{1}{2}}\,\|v\|)\,\|e_\delta(t)\| \;+\; \tfrac{L}{4\sqrt{\varepsilon(t)}}\,\|e_\delta(t)\|^2 \\
&\qquad + r(t) + \tfrac{\delta}{2\sqrt{\varepsilon(t)}}\Big)\,\|e_\delta(t)\|\,.
\end{aligned}$$

To derive (2.6), one uses the formula

$$\frac{d}{dt}\left(\frac{\|e_\delta(t)\|}{r(t)}\right) = \frac{\frac{d}{dt}(\|e_\delta(t)\|^2)}{2\|e_\delta(t)\|\,r(t)} + \frac{-\dot{r}(t)}{r(t)}\,\frac{\|e_\delta(t)\|}{r(t)}\,, \tag{2.12}$$

the relation

$$\begin{aligned}
\frac{-\dot{r}(t)}{r(t)} &= -\frac{\int_0^{m+0}\frac{d}{dt}\left(\frac{\varepsilon(t)}{\varepsilon(t)+\lambda}\right)^2\,d\|E_\lambda(u_0 - y)\|^2}{2\int_0^{m+0}\left(\frac{\varepsilon(t)}{\varepsilon(t)+\lambda}\right)^2\,d\|E_\lambda(u_0 - y)\|^2} \\
&= -\frac{\dot{\varepsilon}(t)}{\varepsilon(t)}\frac{\int_0^{m+0}\frac{\lambda}{\lambda+\varepsilon(t)}\left(\frac{\varepsilon(t)}{\varepsilon(t)+\lambda}\right)^2\,d\|E_\lambda(u_0 - y)\|^2}{\int_0^{m+0}\left(\frac{\varepsilon(t)}{\varepsilon(t)+\lambda}\right)^2\,d\|E_\lambda(u_0 - y)\|^2} \\
&\begin{cases}\le -\frac{\dot{\varepsilon}(t)}{\varepsilon(t)} \le c_\varepsilon \\ \ge 0\end{cases}
\end{aligned} \tag{2.13}$$

with $m := \|\mathcal{A}^*\mathcal{A}\|$, and the decay condition (1.6) on $\varepsilon$.

Lemma 2.2 is proved. $\square$

**Lemma 2.3.** *Assume that conditions (1.5) and (1.6) on $\varepsilon$ and (1.10) and (1.12) on $F$ are satisfied and that $u_\delta(t)$ the solution to (1.18), exists, is unique, and lies in $B(y,\rho)$. Moreover, let either the source condition (1.7) with $\nu \le \frac{1}{2}$ or (1.8) hold. Then the following differential inequality holds*

$$\begin{aligned}\dot{\psi}_\delta(t) \le &-(1-2C_R\|u_0-y\|-c_\varepsilon)\psi_\delta(t)\\ &+\frac{C_R}{2}C_0\|v\|\psi_\delta(t)^2\\ &+1+\frac{\delta}{2\sqrt{\varepsilon(t)}r(t)}\end{aligned} \tag{2.14}$$

*where $C_0$ is as in (1.15), and*

$$\psi_\delta(t) = \frac{\|u_\delta(t)-y\|}{r(t)}\,,$$

$$\begin{aligned}r(t) &= \|\varepsilon(t)(F'(y)^*F'(y)+\varepsilon(t)I)^{-1}(u_0-y)\|\\ &\le \overline{r}(t) := \begin{cases}\nu^\nu(1-\nu)^{1-\nu}\varepsilon(t)^\nu\|v\| & \textit{if (1.7) holds}\\ \gamma_1(p)\left(-\ln\left(\frac{1}{e\varepsilon(0)}\varepsilon(t)\right)\right)^{-p}\|v\| & \textit{if (1.8) holds}\end{cases}\end{aligned} \tag{2.15}$$

*with $\gamma_1(p)$ defined as in Lemma 2.1.*

**Proof of Lemma 2.3.** Starting with formula (2.8) in the proof of Lemma 2.2, we use instead of (2.9) and (2.10) the following relations:

$$\begin{aligned}&A(t)e_\delta(t)+F(y)-F(u_\delta(t)) =\\ &\qquad A(t)\int_0^1[I-R(y+\theta e_\delta(t),u_\delta(t))]d\theta e_\delta(t)\\ &\text{with } \|\int_0^1[I-R(y+\theta e_\delta(t),u_\delta(t))]d\theta\| \le \frac{C_R}{2}\|e_\delta(t)\|\,,\end{aligned} \tag{2.16}$$

and

$$\begin{aligned}&\mathcal{A}^*-A(t)^* = [I-R(u_\delta(t),y)^*]\mathcal{A}^*\,,\\ &\mathcal{A}-A(t) = A(t)[R(y,u_\delta(t)]-I]\,,\\ &\text{with } \|I-R(u_\delta(t),y)^*\| \le C_R\|e_\delta(t)\|\,,\\ &\text{and } \|R(y,u_\delta(t)]-I\| \le C_R\|e_\delta(t)\|\,.\end{aligned} \tag{2.17}$$

This yields

$$\tfrac{d}{dt}(\|e_\delta(t)\|^2) \le 2\Big( -(1-2C_R\|u_0-y\|)\|e_\delta(t)\|^2 + \tfrac{C_R}{2}\|e_\delta(t)\|^3 \\ +r(t) + \tfrac{\delta}{2\sqrt{\varepsilon(t)}}\Big).$$

The rest of the argument is analogous to the one in the proof of Lemma 2.2.

Lemma 2.3 is proved. □

**Lemma 2.4.** *Assume that conditions (1.5) and (1.6) on $\varepsilon$ and (1.10) and (1.13) on $F$ are satisfied and that $u_\delta(t)$ the solution to (1.18) exists, is unique, and lies in $B(y,\rho)$. Moreover, let either the source condition (1.7) with $\nu \le \frac{1}{2}$ or (1.8) hold. Then the following differential inequalities hold:*

$$\begin{aligned}\dot\psi_\delta(t) \le &-(1-2c_\varepsilon)\psi_\delta(t) + c_R(1+c_R)\tilde\psi_\delta(t)\\ &+ \max\{1, c_R(1+c_R)\} + \frac{\delta}{2(\sqrt{\varepsilon(t)}r(t)+\tilde r(t))}\,,\end{aligned} \tag{2.18}$$

$$\begin{aligned}\dot{\tilde\psi}_\delta(t) \le &-(1-c_R(1+c_R)^2-c_\varepsilon)\tilde\psi_\delta(t)\\ &+1+2c_R(1+c_R)^2+(1+c_R)\frac{\delta}{\tilde r(t)}\,,\end{aligned} \tag{2.19}$$

*where*

$$\psi_\delta(t) = \frac{\|u_\delta(t)-y\|}{r(t)+\tilde r(t)/\sqrt{\varepsilon(t)}}\,,$$

$$\tilde\psi_\delta(t) = \frac{\|F'(y)(u_\delta(t)-y)\|}{\tilde r(t)}\,,$$

*$r(t)$ is defined in (2.15),*

$$\begin{aligned}\tilde r(t) &= \|\varepsilon(t)F'(y)(F'(y)^*F'(y)+\varepsilon(t)I)^{-1}(u_0-y)\|\\ &\le \bar{\tilde r} := \begin{cases} (\frac{1}{2}+\nu)^{\frac{1}{2}+\nu}(\frac{1}{2}-\nu)^{\frac{1}{2}-\nu}\varepsilon(t)^{\nu+\frac{1}{2}}\|v\| & \text{if (1.7) holds}\\ \gamma_2(p)\sqrt{\varepsilon(t)}\left(-\ln\left(\frac{1}{e\varepsilon(0)}\varepsilon(t)\right)\right)^{-p}\|v\| & \text{if (1.8) holds}\end{cases}\end{aligned} \tag{2.20}$$

*with $\gamma_1(p)$ and $\gamma_2(p)$ defined as in Lemma 2.1.*

**Proof of Lemma 2.4.** Here instead of (2.9) and (2.10) we have

$$A(t)e_\delta(t) + F(y) - F(u_\delta(t)) = \int_0^1 [I - R(y+\theta e_\delta(t), u_\delta(t))]d\theta R(u_\delta(t), y)\mathcal{A}e_\delta(t) \tag{2.21}$$

with $\| \int_0^1 [I - R(y+\theta e_\delta(t), u_\delta(t))]d\theta R(u_\delta(t), y)\| \le c_R(1+c_R)$,

and

$$\begin{aligned} &\mathcal{A}^* - A(t)^* = A(t)^*[R(y, u_\delta(t))^* - I]\,,\\ &\mathcal{A} - A(t) = [I - R(u_\delta(t), y]\mathcal{A}\\ &\text{with } \|R(y, u_\delta(t))^* - I\| \le c_R\,,\\ &\text{and } \|I - R(u_\delta(t), y\| \le c_R\,, \end{aligned} \tag{2.22}$$

and hence

$$\begin{aligned} \tfrac{d}{dt}(\|e_\delta(t)\|^2) \le 2\Big( &- \|e_\delta(t)\|^2 + c_R(1+c_R)\tfrac{\|\mathcal{A}e_\delta(t)\|}{2\sqrt{\varepsilon(t)}}\|e_\delta(t)\|\\ &+ r(t) + c_R(1+c_R)\tfrac{\tilde{r}(t)}{\sqrt{\varepsilon(t)}} + \tfrac{\delta}{2\sqrt{\varepsilon(t)}}\Big)\,. \end{aligned}$$

To derive a differential inequality for $\frac{\|\mathcal{A}e_\delta(t)\|}{\tilde{r}(t)}$, we apply $\mathcal{A}$ on both sides of (2.8) and use (1.13) to obtain:

$$\begin{aligned} \dot{\mathcal{A}}e_\delta(t) = -\mathcal{A}e_\delta(t) &+ R(y, u_\delta(t))A(t)T_\varepsilon(u)^{-1}A(t)^*\\ &\qquad (A(t)e_\delta(t) + F(y) - F(u_\delta(t)))\\ &+ R(y, u_\delta(t))A(t)T_\varepsilon(u)^{-1}A(t)^*(f_\delta - f) + \varepsilon(t)\mathcal{A}\mathcal{T}_\varepsilon^{-1}(u_0 - y)\\ &+ \varepsilon(t)R(y, u_\delta(t))A(t)T_\varepsilon(u)^{-1}([\mathcal{A}^* - A(t)^*]\mathcal{A}\\ &\qquad + A(t)^*[\mathcal{A} - A(t)])\mathcal{T}_\varepsilon^{-1}(u_0 - y)\,, \end{aligned}$$

with $\|R(y, u_\delta(t))\| \le 1 + c_R$. Using (2.21) and (2.22) one gets:

$$\begin{aligned} \tfrac{d}{dt}(\|\mathcal{A}e_\delta(t)\|^2) \le 2\Big( &- (1 - c_R(1+c_R)^2)\|\mathcal{A}e_\delta(t)\|^2\\ &+ (1 + 2c_R(1+c_R)^2)\tilde{r}(t) + (1+c_R)\delta\Big)\,. \end{aligned}$$

If instead of (2.12) and (2.13) we use

$$\frac{d}{dt}\left(\frac{\|\mathcal{A}e_\delta(t)\|}{\tilde{r}(t)}\right) = \frac{\frac{d}{dt}(\|\mathcal{A}e_\delta(t)\|^2)}{2\|\mathcal{A}e_\delta(t)\|\tilde{r}(t)} + \underbrace{\frac{-\dot{\tilde{r}}(t)}{\tilde{r}(t)}}_{\in[0,c_\varepsilon]}\frac{\|\mathcal{A}e_\delta(t)\|}{\tilde{r}(t)}$$

and

$$\begin{aligned}\frac{d}{dt}\Big(\frac{\|e_\delta(t)\|}{r(t)+\tilde r(t)/\sqrt{\varepsilon(t)}}\Big) = &\frac{\frac{d}{dt}(\|e_\delta(t)\|^2)}{2\,\|e_\delta(t)\|\,(r(t)+\tilde r(t)/\sqrt{\varepsilon(t)})}\\ &+\Bigg(\underbrace{\frac{-\dot r(t)}{r(t)+\tilde r(t)/\sqrt{\varepsilon(t)}}+\frac{-\dot{\tilde r}(t)}{\sqrt{\varepsilon(t)}r(t)+\tilde r(t)}}_{\in[0,2c_\varepsilon]}\\ &+\underbrace{\frac{\tilde r(t)\dot\varepsilon(t)}{2(\sqrt{\varepsilon(t)}r(t)+\tilde r(t))\varepsilon(t)}}_{\le 0}\Bigg)\frac{\|e_\delta(t)\|}{r(t)+\tilde r(t)/\sqrt{\varepsilon(t)}}\,,\end{aligned}$$

then the rest of the argument follows as in the proof of Lemma 2.2.

Lemma 2.4 is proved. □

## 3. PROOF OF THE MAIN RESULTS

**Proof of Theorem 1.1.** To show that the solution $u(t)$ of (1.4) does not leave the ball $B(y,\rho)$, assume the contrary: there exists a $t_1 \in [0,\infty)$ such that $u(t)$ intersects the boundary of $B(y,\rho)$ at $t=t_1$ for the first time:

$$\|u(t_1)-y\| = \rho > \|u(t)-y\| \quad \text{for all} \quad 0 \le t < t_1\,. \tag{3.1}$$

In cases (i) or (ii) we define

$$\psi(t) := \frac{\|u(t)-y\|}{r(t)}$$

and have from Lemma 2.2 or 2.3, respectively, a differential inequality of the form

$$\dot\psi(t) \le C_1 - C_2\psi(t) + C_3\psi^2(t) \quad \text{for all} \quad 0 \le t < t_1\,,$$

where $C_1$, $C_2$ and $C_3$ are positive constants (namely those specified in the statement of the theorem). Since we assume (1.14) we can define

$$\kappa(t) := \kappa_1 + \frac{(\kappa_2-\kappa_1)(\kappa_0-\kappa_1)}{(\kappa_2-\kappa_0)e^{t\sqrt{C_2^2-4C_1C_3}}+\kappa_0-\kappa_1}\,, \tag{3.2}$$

where $\kappa_1$ and $\kappa_2$ solve the scalar quadratic equation $C_1 - C_2\kappa + C_3\kappa^2 = 0$:

$$\kappa_1 := \frac{2C_1}{C_2 + \sqrt{C_2^2 - 4C_1C_3}}\,, \quad \kappa_2 := \frac{C_2 + \sqrt{C_2^2 - 4C_1C_3}}{2C_3}\,, \tag{3.3}$$

and $\kappa_0$ is assumed to lie bewteen $\kappa_1$ and $\kappa_2$:

$$\kappa_1 \le \kappa_0 < \kappa_2\,.$$

By separation of variables one checks that the function $\kappa$ solves the problem:

$$\dot{\kappa}(t) = C_1 - C_2\kappa(t) + C_3\kappa^2(t) \quad \text{for all} \quad t \ge 0\,, \quad \kappa(0) = \kappa_0$$

By the third assumption in (1.14) one may define: $\kappa_0 := \max\{\psi(0), \kappa_1\}$. Using the inequality $\dot{\kappa} \le 0$, one gets:

$$\psi(t) \le \kappa(t) \le \kappa(0) = \max\{\kappa(0), \kappa_1\} \quad \text{for all} \quad 0 \le t < t_1\,. \tag{3.4}$$

Hence, by the monotonicity of $r$ (see (2.13)) and by assumption (1.14), one obtains:

$$\|u(t) - y\| \le \max\{\frac{\|u_0 - y\|}{r(0)}, \kappa_1\}r(0) < \rho \quad \text{for all} \quad 0 \le t < t_1\,,$$

which, for $t \to t_1$, contradicts (3.1) and therefore proves that $u(t)$ remains in $B(y, \rho)$ for all $t > 0$. Moreover, (3.4) implies (1.17).

Consider the case (iii). Assuming again (3.1) for some $t_1 > 0$, we have from Lemma 2.4 a differential inequality

$$\dot{\tilde{\psi}}(t) \le \tilde{C}_1 - \tilde{C}_2\tilde{\psi}(t) \le \max\{\tilde{C}_1, \tilde{\psi}(0)\tilde{C}_2\} - \tilde{C}_2\tilde{\psi}(t) \quad \text{for all} \quad 0 \le t < t_1$$

with

$$\tilde{\psi}(t) := \frac{\|\mathcal{A}e(t)\|}{\tilde{r}(t)}\,.$$

With $\tilde{\kappa}(t) := \max\{\frac{\tilde{C}_1}{\tilde{C}_2}, \tilde{\psi}(0)\}$ solving $\dot{\tilde{\kappa}}(t) = \max\{\tilde{C}_1, \tilde{\psi}(0)\tilde{C}_2\} - \tilde{C}_2\tilde{\kappa}(t)$ for all $0 < t$, and using the inequality $\tilde{\kappa}(0) \ge \tilde{\psi}(0)$ one can conclude

$$\tilde{\psi}(t) \le \tilde{\kappa}(t) = \max\{\frac{\tilde{C}_1}{\tilde{C}_2}, \tilde{\psi}(0)\} \quad \text{for all} \quad 0 \le t < t_1\,.$$

Inserting this into (2.18) one gets a differential inequality

$$\dot{\psi}(t) \leq C_1 - C_2\psi(t) \quad \text{for all} \quad 0 \leq t < t_1$$

with

$$\psi(t) := \frac{\|u(t) - y\|}{r(t) + \tilde{r}(t)/\sqrt{\varepsilon(t)}}$$

yielding as above

$$\psi(t) \leq \max\{\frac{C_1}{C_2}, \psi(0)\} \quad \text{for all} \quad 0 \leq t < t_1 \,.$$

Assumption (1.16) leads to a contradiction to (3.1). Therefore $u(t) \in B(y, \rho)$ for all $t \geq 0$ and (1.17) holds.

Theorem 1.1 is proved. □

**Proof of Corollary 1.3.** Replacing $r(t)$ by its upper esimate $\overline{r}(t)$ in (2.15), one can proceed as in the proofs of Lemmas 2.2, 2.3 to obtain the differential inequality for $\psi_\delta$ replaced by

$$\overline{\psi}_\delta(t) = \frac{\|u_\delta(t) - y\|}{\overline{r}(t)}$$

and $r(t)$ replaced by $\overline{r}(t)$. From (1.20) or (1.21) and the strict monotonicity of $\varepsilon(t)$, one gets the inequality

$$\frac{\delta}{2\sqrt{\varepsilon(t)}\overline{r}(t)} \leq \frac{\overline{C}}{\tau} \quad \text{for all} \quad 0 < t < t_\delta$$

which holds for some constant $\overline{C} > 0$. Thus the function $\overline{\psi}_\delta$ in cases (i) and (ii) satisfies the differential inequality

$$\dot{\overline{\psi}}_\delta(t) \leq \overline{C}_1 - C_2\overline{\psi}_\delta(t) + C_3\overline{\psi}_\delta^2(t) \quad \text{for all} \quad 0 < t < t_\delta \,,$$

with $\overline{C}_1 = C_1 + \frac{\overline{C}}{\tau}$. By making $\tau$ sufficiently large and therefore $\frac{\overline{C}}{\tau}$ small, conditions (1.14) with $C_1$ replaced by $\overline{C}_1$ can be satisfied and one concludes as in the proof of Theorem 1.1 that for all times $t < t_\delta$

$$\|e_\delta(t)\| \leq \max\{\overline{\psi}_\delta(0), \kappa_1\}\overline{r}(t) \,.$$

Letting $t$ tend to $t_\delta$, one gets:

$$\|e_\delta(t_\delta)\| \le C\varepsilon(t_\delta)^\nu = C\tau^{\frac{2\nu}{2\nu+1}}\,\delta^{\frac{2\nu}{2\nu+1}}$$

in case of (1.7), or

$$\|e_\delta(t_\delta)\| \le C(-\ln(\frac{1}{\varepsilon(0)e}\varepsilon(t_\delta)))^{-p} = C(-\ln(\frac{1}{\sqrt{\varepsilon(0)}e}\,\tau\delta))^{-p}$$

in case of (1.8), with some constant $C > 0$. Convergence (1.22) in the situation when $\nu = 0$ and $v = u_0 - y$ in the definition of $\overline{r}$, $\overline{\psi}_\delta$ and when we only assume $u_0 - y \in N(\mathcal{A})^\perp$, follows directly from the slightly sharper differential inequality

$$\dot{\overline{\psi}}_\delta(t) \le C_1(\delta) - C_2\overline{\psi}_\delta(t) + C_3\overline{\psi}_\delta^2(t) \quad \text{for all} \quad 0 < t < t_\delta\,,$$

where

$$C_1(\delta) = \frac{r(t_\delta)}{\|u_0 - y\|} + \frac{\delta}{2\sqrt{\varepsilon(t_\delta)}\,\|u_0 - y\|}\,.$$

By Lemma 2.1 and formula (1.19), $C_1(\delta) \to 0$ as $\delta \to 0$. Namely, as in (3.2), (3.3), (3.4) in the proof of Theorem 1.1, one gets, replacing $C_1$ by $C_1(\delta)$, and letting $t$ tend to $t_\delta$

$$\overline{\psi}_\delta(t_\delta) \le \kappa_1(\delta) + \frac{(\kappa_2(\delta) - \kappa_1(\delta))(\kappa_0 - \kappa_1(\delta))}{(\kappa_2(\delta) - \kappa_0)e^{t_\delta\sqrt{C_2^2 - 4C_1(\delta)C_3}} + \kappa_0 - \kappa_1(\delta)} \tag{3.5}$$

with

$$\kappa_1(\delta) = \frac{2C_1(\delta)}{C_2 + \sqrt{C_2^2 - 4C_1(\delta)C_3}}\,, \quad \kappa_2(\delta) = \frac{C_2 + \sqrt{C_2^2 - 4C_1(\delta)C_3}}{2C_3}\,.$$

Now by the inequality $C_1(\delta) \le \overline{C}_1$ and assumptions (1.14) which are valid with $C_1$ replaced by $\overline{C}_1$, because $\tau$ is chosen sufficiently large, one gets for the two terms on the right-hand side of (3.5)

$$\kappa_1(\delta) \le CC_1(\delta)$$

and

$$\frac{(\kappa_2(\delta) - \kappa_1(\delta))(\kappa_0 - \kappa_1(\delta))}{(\kappa_2(\delta) - \kappa_0)e^{t_\delta\sqrt{C_2^2 - 4C_1(\delta)C_3}} + \kappa_0 - \kappa_1(\delta)} \le \frac{C}{e^{t_\delta\sqrt{C_2^2 - 4\overline{C}_1C_3}} - C}$$

for some constant $C > 0$, so that by (1.19) these terms both go to zero as $\delta \to 0$. This and the relation $\|u_\delta(t_\delta) - y\| = \|u_0 - y\| \overline{\psi}_\delta(t_\delta)$ imply (1.22).

Analogously the proof in (iii) of Theorem 1.1 can be modified to yield (1.22), (1.23).

Corollary 1.3 is proved. $\square$

**Proof of Theorem 1.4.** As before assume that $u_\delta(t)$ leaves $B(y,\rho)$ at $t = t_1 < t_\delta$ for the first time, i.e., (3.1). The nonlinearity condition (1.13) implies

$$\begin{aligned}(1 - c_R)\, \|F'(u)(\bar{u} - u)\| &\le \|F(\bar{u}) - F(u)\| \\ &\le (1 + c_R)\, \|F'(u)(\bar{u} - u)\|\,, \quad u, \bar{u} \in B(y,\rho)\end{aligned} \tag{3.6}$$

and, by (1.25) and (1.3), one gets

$$\tau\delta < \|F(u_\delta(t)) - f_\delta\| \le (1 + c_R)\, \|\mathcal{A} e_\delta(t)\| + \delta\,,$$

so that

$$(\tau - 1)\delta < (1 + c_R)\, \|\mathcal{A} e_\delta(t)\| \quad \text{for all} \quad 0 < t < t_1 \tag{3.7}$$

Inserting this into the last term of (2.19), one gets:

$$\frac{d}{dt}\left(\frac{\|\mathcal{A} e_\delta(t)\|}{\tilde{r}(t)}\right) \le \tilde{C}_1 - \overline{\tilde{C}}_2 \left(\frac{\|\mathcal{A} e_\delta(t)\|}{\tilde{r}(t)}\right) \quad \text{for all} \quad 0 < t < t_1\,,$$

with $\tilde{C}_1 = 1 + 2c_R(1 + c_R)^2$, $\overline{\tilde{C}}_2 = 1 - c_R(1 + c_R)^2 - c_\varepsilon - \frac{(1+c_R)^2}{\tau - 1}$. As in the proof of Theorem 1.1 this yields

$$\frac{\|\mathcal{A} e_\delta(t)\|}{\tilde{r}(t)} \le \max\{\frac{\tilde{C}_1}{\overline{\tilde{C}}_2}, \frac{\|\mathcal{A} e_\delta(0)\|}{\tilde{r}(0)}\} \quad \text{for all} \quad 0 < t < t_1$$

and therefore by (2.18) and (3.7)

$$\frac{d}{dt}\left(\frac{\|e_\delta(t)\|}{r(t) + \tilde{r}(t)/\sqrt{\varepsilon(t)}}\right) \le \overline{C}_1 - C_2 \left(\frac{\|e_\delta(t)\|}{r(t) + \tilde{r}(t)/\sqrt{\varepsilon(t)}}\right) \text{ for all } 0 < t < t_1\,.$$

This and conditions (1.16) imply

$$\|e_\delta(t)\| \le \max\{\frac{\overline{C}_1}{C_2} r(0), \|u_0 - y\|\} < \rho\,,$$

hence contradicting (3.1). Therefore

$$\begin{aligned} \|e_\delta(t)\| &\le C(r(t)+\tilde r(t)/\sqrt{\varepsilon(t)})\,, \\ \|\mathcal{A}e_\delta(t)\| &\le C\tilde r(t)\,, \end{aligned} \quad \text{for all} \quad 0<t\le t_\delta\,, \tag{3.8}$$

with some constant $C>0$ independent of $\delta$ and $t$. From (3.7) and (3.8) one gets:

$$\delta \le \frac{1+c_R}{\tau-1}C\tilde r(t) \le \tilde C\varepsilon(t)^{\nu+\frac12} \quad \text{for all} \quad 0<t<t_\delta\,. \tag{3.9}$$

On the other hand one can use (1.24) to derive an estimate of $\delta$ from below of the form

$$\delta \ge c\tilde r(t_\delta)$$

with some constant $c>0$. To do this, one derives analogously to (2.19), that

$$\begin{aligned} \frac{d}{dt}\left(\frac{\|\mathcal{A}e_\delta(t)\|}{\tilde r(t)}\right) &\ge -(1+c_R(1+c_R)^2)\left(\frac{\|\mathcal{A}e_\delta(t)\|}{\tilde r(t)}\right) \\ &\qquad +1-2c_R(1+c_R)^2-(1+c_R)\frac{\delta}{\tilde r(t)} \\ &\ge \hat C_1-\hat C_2\left(\frac{\|\mathcal{A}e_\delta(t)\|}{\tilde r(t)}\right) \end{aligned}$$

for all $0<t\le t_\delta$, where $\hat C_1 = 1-2c_R(1+c_R)^2$ and $\hat C_2 = 1+c_R(1+c_R)^2+\frac{(1+c_R)^2}{\tau-1}$. Thus,

$$\frac{\|\mathcal{A}e_\delta(t)\|}{\tilde r(t)} \ge \min\{\frac{\hat C_1}{\hat C_2},\frac{\|\mathcal{A}e_\delta(0)\|}{\tilde r(0)}\} \quad \text{for all} \quad 0<t<t_\delta\,,$$

where the lower bound $\min\{\frac{\hat C_1}{\hat C_2},\frac{\|\mathcal{A}e_\delta(0)\|}{\tilde r(0)}\}$ is strictly positive, due to the assumption $u_0-y\in N(F'(y)^\perp$. Letting $t$ tend to $t_\delta$ here, and using the stopping criterion (1.24), assumption (1.3) and inequalities (3.6), one gets:

$$\tilde r(t_\delta) \le \frac{\tau+1}{1-c_R}\left(\min\{\frac{\hat C_1}{\hat C_2},\frac{\|\mathcal{A}e_\delta(0)\|}{\tilde r(0)}\}\right)^{-1}\delta\,, \tag{3.10}$$

and by the interpolation inequality

$$\|T^a v\| \le \|T^b v\|^{\frac{a}{b}}\,\|v\|^{1-\frac{a}{b}} \quad \text{for } 0<a<b<\infty\,,$$

which holds for nonnegative definite selfadjoint, not necessarily bounded operators $T$, one obtains:

$$r(t_\delta) \leq \tilde{r}(t_\delta)^{\frac{2\nu}{2\nu+1}} \|v\|^{\frac{1}{2\nu+1}} \leq \tilde{C}\delta^{\frac{2\nu}{2\nu+1}} \tag{3.11}$$

with some constant $\tilde{C} > 0$. Taking $t = t_\delta$ in (3.8), and using (3.9), (3.10), and (3.11), one gets:

$$\begin{aligned} \|e_\delta(t_\delta)\| &\leq C(r(t_\delta) + \tfrac{\tilde{r}(t_\delta)}{\sqrt{\varepsilon(t_\delta)}}) \\ &\leq C(\delta^{\frac{2\nu}{2\nu+1}} + \tfrac{\delta}{\sqrt{\varepsilon(t_\delta)}}) \\ &\leq C\delta^{\frac{2\nu}{2\nu+1}}(1 + \tfrac{\delta^{\frac{1}{2\nu+1}}}{\sqrt{\varepsilon(t_\delta)}}) = O(\delta^{\frac{2\nu}{2\nu+1}}), \end{aligned} \tag{3.12}$$

(with constants $C$ independent of $\delta$ possibly taking different values) which is assertion (1.29).

If $\nu = 0$, i.e., no regularity of $u_0 - y$ is assumed, then by (2.5) and the strict monotonicity of $\varepsilon(t)$, one concludes from (3.11) that $t_\delta \to \infty$ as $\delta \to 0$. This implies $\varepsilon(t_\delta) \to 0$ as $\delta \to 0$, so that from the first line in (3.12) it follows by (2.4) that $\|e_\delta(t_\delta)\| \to 0$ as $\delta \to 0$.

Theorem 1.4 is proved. $\square$

# References


[1] R.G. Airapetyan, A.G. Ramm, A.B. Smirnova, *Continuous analog of the Gauss-Newton method*, Math. Models and Methods in Appl. Sci. 9 (1999), 463-474.

[2] R.G. Airapetyan, A.G. Ramm, *Dynamical systems and discrete methods for solving nonlinear ill-posed problems*, Appl.Math.Reviews, vol. 1, G. Anastassiou ed., World Sci. Publishers, 2000, pp.491-536.

[3] R.G. Airapetyan, A.G. Ramm, A.B. Smirnova, *Continuous regularization of nonlinear ill-posed problems*, in the book "Operator theory and applications", A.G.Ramm, P.N.Shivakumar, A.V.Strauss eds., Amer. Math. Soc., Fields Institute Communications, Providence,RI, 2000, pp. 111-138.

[4] B. Blaschke(-Kaltenbacher), A. Neubauer, and O. Scherzer, *On convergence rates for the iteratively regularized Gauß–Newton method*, IMA J. Numer. Anal. 17, (1997), 421–436.

[5] P. Deuflhard, H.W. Engl, and O. Scherzer, *A convergence analysis of iterative methods for the solution of nonlinear ill-posed problems under affinely invariant conditions*, Inverse Problems 14 (1998), 1081–1106.

[6] H.W. Engl, M. Hanke, A. Neubauer, *Regularization of Inverse Problems*, Kluwer, Dordrecht, 1996

[7] M. Hanke, A. Neubauer, and O. Scherzer, *A convergence analysis of the Landweber iteration for nonlinear ill-posed problems*, Numer. Math. 72 (1995), 21–37.

[8] B. Hofmann and O. Scherzer, *Influence factors if ill-posedness for nonlinear problems*, Inverse Problems 10 (1994), 1277–1297.

[9] T. Hohage, *Logarithmic convergence rates of the iteratively regularized Gauß-Newton method for an inverse potential and an inverse scattering problem*, Inverse Problems 13 (1997), 1279-1299.

[10] T. Hohage, *Regularization of exponentially ill-posed problems*, Numer. Funct. Anal. Optim. 21 (2000), 439-464.

[11] B. Kaltenbacher, *On Broyden's method for nonlinear ill–posed problems* Numer. Funct. Anal. Optim. 19 (1998), 807-833.

[12] A.G. Ramm, *Stable solutions of some ill-posed problems*, Math. Meth. in the Appl. Sci. 3, (1981), 336-363.

[13] A.G. Ramm, *Linear ill-posed problems and dynamical systems*, Jour. Math. Anal. Appl. 258, (2001)

[14] A.G. Ramm, A.B. Smirnova, *A numerical method for solving nonlinear ill-posed problems*, Nonlinear Funct. Anal. and Optimiz. 20 (1999), 317-332.

[15] A.G. Ramm, A.B. Smirnova, *On stable numerical differentiation*, Mathem. of Computation 70, (2001), 1131-1153.

[16] A.G. Ramm, A.B. Smirnova, *Continuous regularized Gauss-Newton-type algorithm for nonlinear ill-posed equations with simultaneous updates of inverse derivative*, (submitted).

[17] A.G. Ramm, A.B. Smirnova, A. Favini, *Continuous modified Newton's-type method for nonlinear operator equations*, (submitted).